

# Multi-Camera Occlusion and Sudden-Appearance-Change Detection Using Hidden Markovian Chains

Xudong Ma

Pattern Technology Lab LLC, U.S.A.

Email: xma@ieee.org

**Abstract**—This paper was originally submitted to Xnova as a response to a Request for Invention (RFI) on new event monitoring methods. In this paper, a new object tracking algorithm using multiple cameras for surveillance applications is proposed. The proposed system can detect sudden-appearance-changes and occlusions using a hidden Markovian statistical model. The experimental results confirm that our system detect the sudden-appearance changes and occlusions reliably.

**Index Terms**—Computer Vision, Object Tracking, Hidden Markovian Model, Bayesian Estimation, Surveillance

## I. INTRODUCTION

In surveillance applications, the ability of long-term tracking a certain person or object from security cameras is highly desirable. Usually the security personnel can label a suspicious person in the video. The tracking system can then track this suspicious person in the videos without further human inputs for minutes or hours.

One situation that the visual tracking systems for the surveillance applications may deal with frequently is the sudden-appearance-changes of the being tracked person or object. For example, the suspicious person may take off his/her jacket, pull up his/her hood, or abandon some luggage, in order to fool the surveillance system. Such sudden-appearance-changes are suspicious activities and usually should be reported to the security personnel in real-time.

Unfortunately, it is usually difficult for computer vision algorithms to distinguish between a sudden-appearance-change to occlusions. For many surveillance applications, the surveillance scenes are usually crowded with many people. Thus, occlusions

may happen very frequently. Without the ability of distinguishing between sudden-appearance-changes to occlusions, the visual tracking systems may generate a large number of false alarms.

Detecting occlusions correctly is also important for enhancing the reliability of visual tracking algorithms. It is well-known that for visual tracking, there exists a so called “high-adaptability-to-drifting-resistance trade-off” problem. The problem relates to how much we should update our models for the being-tracked person or object on-the-fly. If we do not update the models much, then in the case that the appearance of the being tracked person changes rapidly, there is a high risk of tracking loss. If we always update the models very rapidly, then in the case of occlusions, the model may be tuned to occluders. And these wrongly tuned models may result in the so called drifting. That is, the tracking algorithm may start to track the occluders instead. In Fig. 2 of section IV, we actually show an example of such drifting phenomena.

In this paper, we propose a new occlusion and appearance-change detection method. The proposed real-time visual tracking system uses multiple surveillance cameras. Initially, the security personnel provides one bounding box of the suspicious person for each video frame sequence. The visual tracking system then tracks the where-about of the suspicious person in real-time. If a sudden-appearance-change of the suspicious person is detected, then the visual tracking system would raise an alarm signal immediately.

Our method uses both generative and discriminative models for the video frame streams. For each camera, one discriminative model is maintained for discriminating the image patches that contain the

being-tracked person to the image patches that do not contain the being-tracked person. Similarly, one generative model is maintained for each camera. In this paper, we use a recently proposed compressive sensing and naive Bayes based classifier in [1] as the discriminative model. We use linear sub-space models as the generative models. That is, we assume that the image patches containing the being-tracked person from several adjacent video frames all are vectors within a certain affine sub-space.

A center component of our method is a hidden Markovian model for the prediction errors of the generative models. That is, whenever a new video frame is received, the new image patch containing the being-tracked person is predicted from the previous such image patches. The hidden Markovian model thus contains a visible part and a hidden part. The visible part contains the observed prediction errors. And the hidden part contains random variables  $O_1(t), O_2(t), \dots, O_N(t)$  and  $S(t)$ . The binary random variable  $O_n(t)$  denotes whether an occlusion has occurred for the  $n$ -th camera at time  $t$ . And the binary random variable  $S(t)$  denotes whether a sudden-appearance-change of the being-tracked person has occurred at time  $t$ .

We assume some parametric probability distributions for the hidden Markovian model. The probabilities of  $O_n(t)$  and  $S(t)$  are estimated from the observed prediction error  $z_n(t)$  by using sequential Bayesian estimation. An alarm signal may be raised, if we detect a high probability of  $S(t) = 1$ , a sudden-appearance-change occurred. The estimated probabilities of  $O_n(t)$  and  $S(t)$  are also used for adjusting the learning rates of the discriminative and generative models. It should be intuitively clear that any appearance-change of the being tracked objects may result in significant prediction errors at the same time at all the cameras, but occlusions usually result in significant prediction errors only at a few cameras. Their probabilities can thus be estimated accordingly.

Please note that the above hidden Markovian statistical model is the centerpiece of the proposed occlusion and sudden-appearance-change detection methods. The statistical model works also well with other discriminative and generative models.

There is a large literature on visual tracking algorithms, such as [2], [3], [4]. We have no intention here to provide a throughout survey on the general visual tracking algorithms. The approaches of using

multiple cameras for visual tracking have become attractive in the recent years, due to the availability of large quantities of low-cost commodity cameras. There are some previous discussions on visual tracking using multiple cameras. In [5] [6], approaches are discussed, where the responsibilities of tracking may be passed from one camera to another camera. In [7], from each camera, a statistical estimation of the location of the being tracked person or object is obtained independently. The independent estimations are then fused into a joint location estimation. In [8], video frames from multiple cameras are projected on a reference frame using homography transforms, such that the signals corresponding to the being tracked person or object may be added constructively.

The rest of the paper is organized as follows. In Section II, we discuss the proposed visual tracking system and the hidden Markovian model. In Section III, we present the sequential Bayesian estimation methods for the hidden Markovian model. Simulation results of the proposed method is provided in Section IV. Finally, some concluding remarks are presented in Section V.

## II. VISUAL TRACKING SYSTEM AND HIDDEN MARKOVIAN MODEL

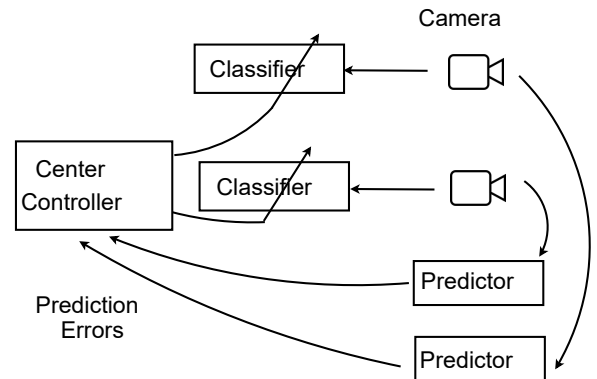


Fig. 1: Block diagram of the visual tracking system

A block diagram of the proposed visual tracking system is shown in Fig. 1. The system uses multiple cameras (only 2 are shown here). The system starts tracking a suspicious person, after the security personnel provides one bounding box of the suspicious person for each camera. For each camera, there is a real-time tracking sub-system as in [1] (shown as classifiers in Fig. 1). All such real-time tracking sub-systems work almost independently, except that

their learning parameters  $\lambda$  are controlled by the center controller.

The learning parameters  $\lambda$  control how much each individual real-time compressive tracking sub-system should update the discriminative model after receiving each new video frame. As the being-tracked person changes his/her pose, orientation etc., the appearance of the person may change smoothly. Thus, each compressive tracking sub-system may update its discriminative model according to each newly observed video frame. The parameter  $\lambda$  is a real number between 0 and 1. If  $\lambda = 1$ , then the compressive tracking sub-system does not update its model. If  $0 < \lambda < 1$ , then the compressive tracking sub-system updates the discriminative model using a combination of past observed video frames and the newly observed video frame.

Our proposed method adjusts the learning parameters  $\lambda$  based on the probabilities of sudden-appearance-change  $\mathbb{P}[S(t)]$  and the probabilities of occlusions  $\mathbb{P}[O_n(t)]$ . Let us assume that we use  $N$  security cameras. Suppose at each time  $t = 1, 2, \dots$ , we receive one video frame  $x_n(t)$  at each camera, where  $1 \leq n \leq N$ . Each tracking sub-system then finds one image patch  $Y_n(t)$  that contains the being-tracked person, where  $Y_n(t)$  is a column vector. That is,  $Y_n(t)$  is the vector obtained by stacking the pixels in the image patch.

For each camera, we use one generative model (shown as predictors in Fig. 1). Each predictor maintains an estimation of an affine subspace  $Space_n(t)$ , such that all the past observed  $Y_n(t-1), Y_n(t-2), \dots$  roughly lie within this affine space. We can then define a prediction error  $Z_n(t)$  as the distance of the newly observed  $Y_n(t)$  to this affine space  $Space_n(t)$ . Note that there exist efficient algorithms for computing the affine space  $Space_n(t)$ , such as in [9].

We may then estimate the probabilities of sudden-appearance-change  $\mathbb{P}[S(t)]$  and the probabilities of occlusions  $\mathbb{P}[O_n(t)]$  from  $Z_n(t)$  based on the following hidden Markovian model. We assume that the probability density function of  $Z_n(t)$

$$\begin{aligned} & \mathbb{P}[Z_n(t)|S(t), O_n(t)] \\ &= \begin{cases} \exp(-x/\mu)/\mu, & \text{if } S(t) = O_n(t) = 0 \\ 1/M, & \text{otherwise} \end{cases} \end{aligned}$$

where,  $S(t) = 1$  indicates that a sudden-appearance-change has occurred at time  $t$  and  $O_n(t) = 1$

indicates that an occlusion has occurred at the  $n$ -th camera at time  $t$ . In other words, if  $S(t) = O_n(t) = 0$ , then the prediction error  $Z_n(t)$  is exponentially distributed. Otherwise, the prediction error  $Z_n(t)$  is uniformly distributed between 0 and  $M$ . We further assume that  $Z_n(t)$  and  $Z_m(t)$  are statistically independent conditioned on  $S(t)$  and  $O_1(t), \dots, O_N(t)$ , if  $n \neq m$ . We assume that each random process  $O_n(1), O_n(2), \dots, O_n(t)$  is a Markovian random process. Similarly, We also assume that the random process  $S(1), S(2), \dots, S(t)$  is Markovian.

We may then use the Bayesian decision methods to detect occlusions at each camera  $O_n(t)$  and the sudden appearance change of the being tracked person or object  $S(t)$  by computing the probabilities

$$\begin{aligned} & \mathbb{P}[S(t) = 1|Z_{1:N}(1:t)] \\ &= \mathbb{P}[S(t) = 1|Z_1(1:t), Z_2(1:t), \dots, Z_N(1:t)] \\ & \mathbb{P}[O_n(t) = 1|Z_{1:N}(1:t)] \\ &= \mathbb{P}[O_n(t) = 1|Z_1(1:t), Z_2(1:t), \dots, Z_N(1:t)] \end{aligned}$$

where  $Z_n(1:t)$  denotes the collection of observed prediction errors  $Z_n(1), Z_n(2), \dots, Z_n(t)$ , and  $Z_{1:N}(1:t)$  denote the collection of variables  $Z_1(1:t), \dots, Z_N(1:t)$ . We show in Section III that these probabilities can be recursively calculated in a very efficient way.

The proposed method raises an alarm signal, whenever the probability  $\mathbb{P}[S(t) = 1|Z_{1:N}(1:t)]$  goes over a certain threshold. The method may also adjust the learning parameters  $\lambda$  of the compressive tracking sub-systems according to the probability  $\mathbb{P}[O_n(t) = 1|Z_{1:N}(1:t)]$ . For example, we may set  $\lambda = 1$ , whenever the probability  $\mathbb{P}[O_n(t) = 1|Z_{1:N}(1:t)]$  is greater than 0.5.

### III. RECURSIVE BAYESIAN ESTIMATION

In this section, we derive a recursive formula for calculating  $\mathbb{P}[S(t), O_{1:N}(t)|Z_{1:N}(1:t)]$  from  $\mathbb{P}[S(t-1), O_{1:N}(t-1)|Z_{1:N}(1:t-1)]$  in Eq. 1, where (a) follows from the Markovian properties of the statistical model.

### IV. EXPERIMENTAL RESULT

We use one of the PETS 2007 (Tenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance) data-sets (available from <http://www.cvg.reading.ac.uk/PETS2007/data.html>). The data-set consists of 12000 video frames for 4

$$\begin{aligned}
\mathbb{P}[S(t), O_{1:N}(t) | Z_{1:N}(1:t)] &= \frac{\mathbb{P}[S(t), O_{1:N}(t), Z_{1:N}(1:t)]}{\mathbb{P}[Z_{1:N}(1:t)]} \propto \mathbb{P}[S(t), O_{1:N}(t), Z_{1:N}(1:t)] \\
&\propto \mathbb{P}[S(t), O_{1:N}(t), Z_{1:N}(t) | Z_{1:N}(1:t-1)] \\
&= \sum_{S(t-1), O_{1:N}(t-1)} \mathbb{P}[S(t), O_{1:N}(t), S(t-1), O_{1:N}(t-1), Z_{1:N}(1:t) | Z_{1:N}(1:t-1)] \\
&= \sum_{S(t-1), O_{1:N}(t-1)} \mathbb{P}[S(t-1), O_{1:N}(t-1), | Z_{1:N}(1:t-1)] \\
&\quad \times \mathbb{P}[S(t), O_{1:N}(t) | S(t-1), O_{1:N}(t-1), Z_{1:N}(1:t-1)] \\
&\quad \times \mathbb{P}[z_{1:N}(t) | S(t), O_{1:N}(t), S(t-1), O_{1:N}(t-1), Z_{1:N}(1:t-1)] \\
&\stackrel{(a)}{=} \sum_{S(t-1), O_{1:N}(t-1)} \mathbb{P}[S(t-1), O_{1:N}(t-1), | Z_{1:N}(1:t-1)] \\
&\quad \times \mathbb{P}[S(t), O_{1:N}(t) | S(t-1), O_{1:N}(t-1)] \mathbb{P}[z_{1:N}(t) | S(t), O_{1:N}(t)]
\end{aligned} \tag{1}$$

cameras, 3000 video frames for each camera. A suspicious person enters the scene at roughly frame 500 and drops and leaves behind his backpack at roughly frame 850.

We observe that the real-time compressive tracking systems with a fixed learning parameter in the prior art [1] do have the drifting problem as shown in Fig. 2. We show in Fig. 3 that such drifting problem can be avoided by the algorithm proposed in this paper. The occlusion event around frame 577 is detected by our algorithm very clearly with the corresponding probability close to 1.

The proposed algorithm also detects the unusual behavior of the being-tracked person at frame 851 (with the corresponding probability higher than 0.9). The proposed algorithm is able to track the whereabouts of the suspicious person as shown in Figs. 4, and 5, where each figure shows the tracking results at one camera.

In all the above experimental results, the tracking results are shown as red bounding boxes. And the frame indexes are labelled in all the images.

## V. CONCLUSION

The paper discusses a visual tracking algorithm to detect sudden-appearance-change and occlusions. By experimental results, we show that the proposed algorithm can reliably detect the sudden-appearance-change and occlusion events. Such reliable estimations can also be used to avoid the drifting problems.

## ACKNOWLEDGMENT

This research was originally submitted to Xinova, LLC by the author in response to a Request for Invention. It is among several submissions that Xinova has chosen to make available to the wider community. The author wishes to thank Xinova, LLC for their funding support of this research. More information about Xinova, LLC is available at [www.xinova.com](http://www.xinova.com).

## REFERENCES

- [1] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *12th European Conference on Computer Vision (ECCV)*, October 2012.
- [2] V. R. D. Comaniciu and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, 2003.
- [3] S. Avidan, "Support vector tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, 2004.
- [4] —, "Ensemble tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, 2007.
- [5] M. Quaritsch, M. Kreuzthaler, B. Rinner, H. Bischof, and B. Strobl, "Autonomous multicamera tracking on embedded smart cameras," *Journal on Embedded Systems*, January 2007.
- [6] Q. Cai and J. Aggarwal, "Tracking human motion in structured environments using a distributed-camera system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, 1999.
- [7] M. Bhuyan, B. Lovell, and A. Bigdeli, "Tracking with multiple cameras for video surveillance," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, December 2007.
- [8] R. Eshel and Y. Moses, "Homography based multiple camera detection and tracking of people in a dense crowd," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008.
- [9] A. Levy and M. Lindenbaum, "Sequential Karhunen-Loeve basis extraction and its applications to images," *IEEE Transactions on Image Processing*, vol. 9, no. 8, 2000.



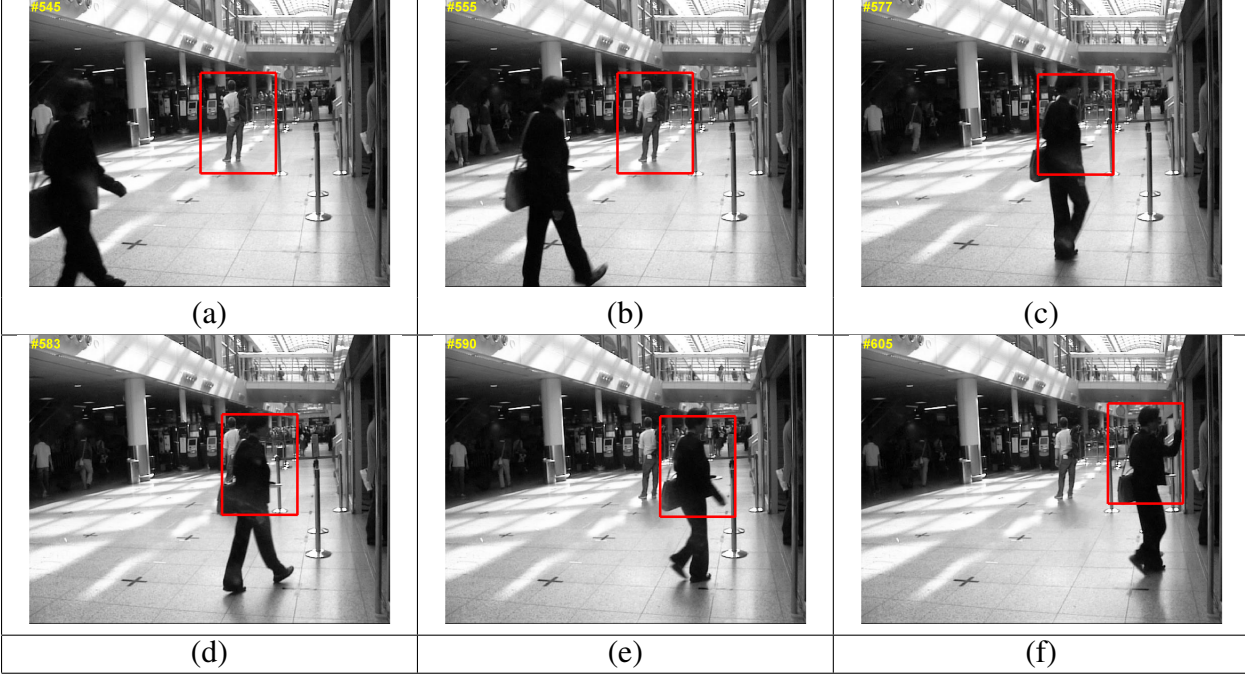


Fig. 2: Drifting problem of the real-time compressive tracking algorithm with a fixed learning parameter  $\lambda = 0.85$ . When the being-tracked person is completely occluded by the passenger at frame 577 (sub-figure (c)), the discriminative model is updated according to the appearance of the passenger. Thus, a tracked loss occurs. (the figure is best viewed in color)

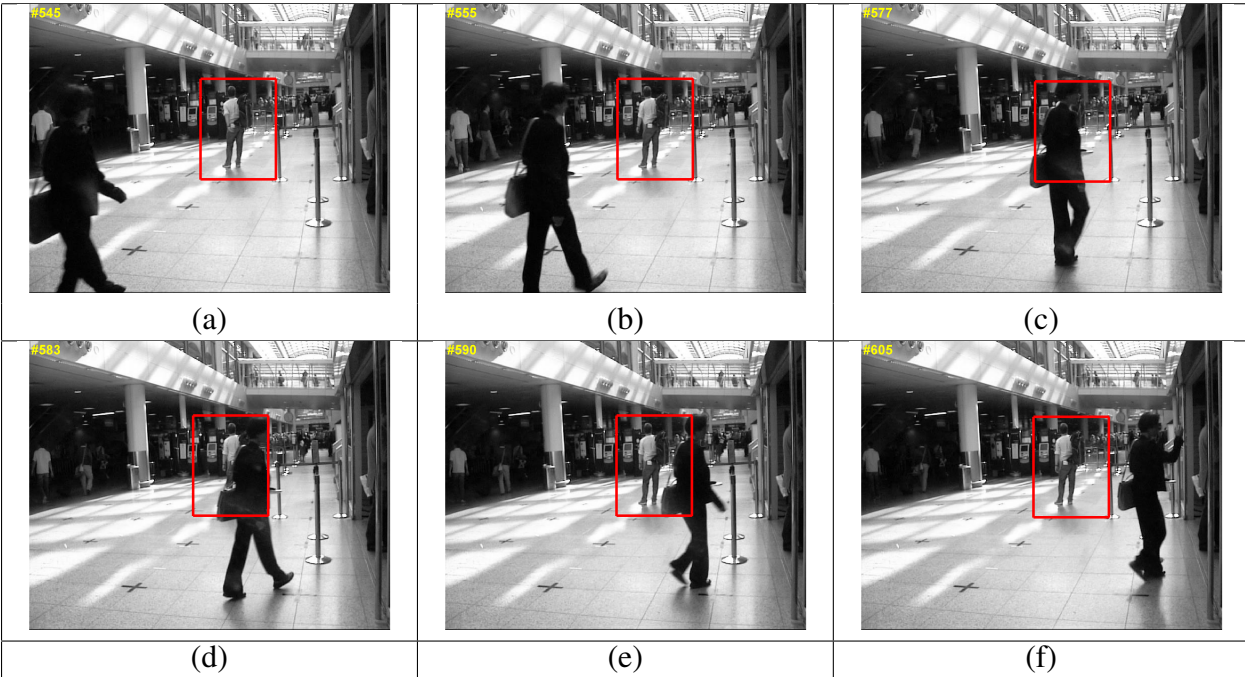


Fig. 3: Drifting is avoided by the proposed algorithm

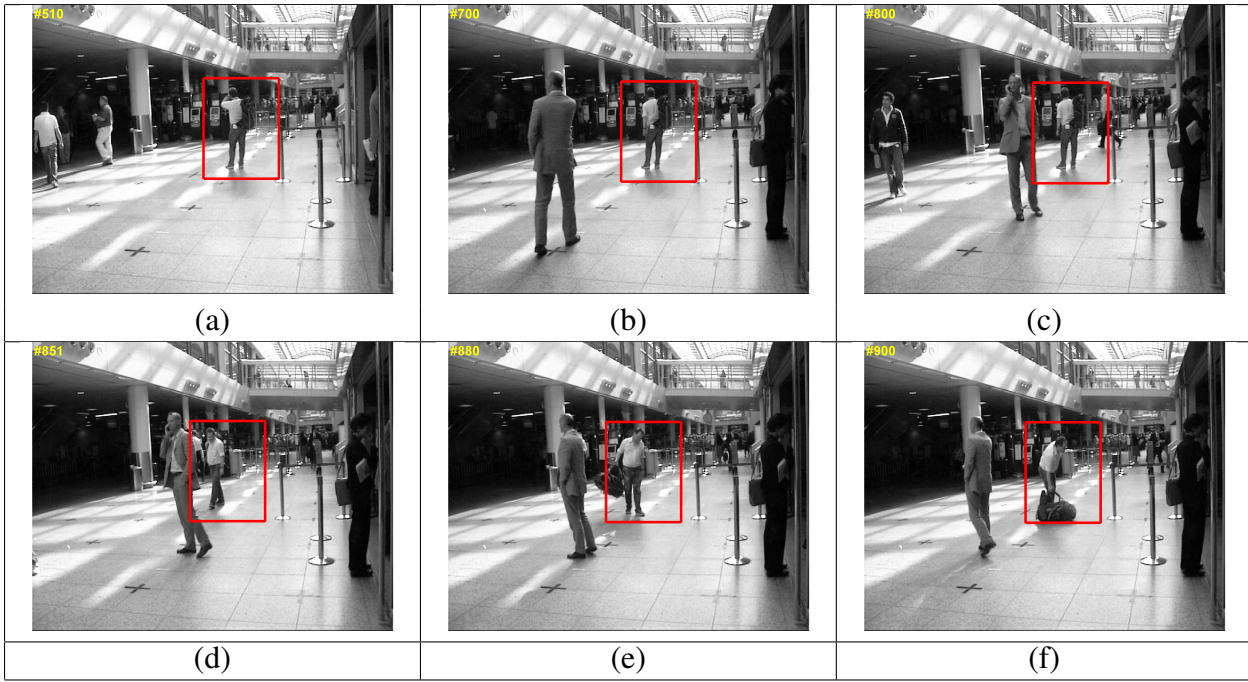


Fig. 4: Detecting sudden-appearance change

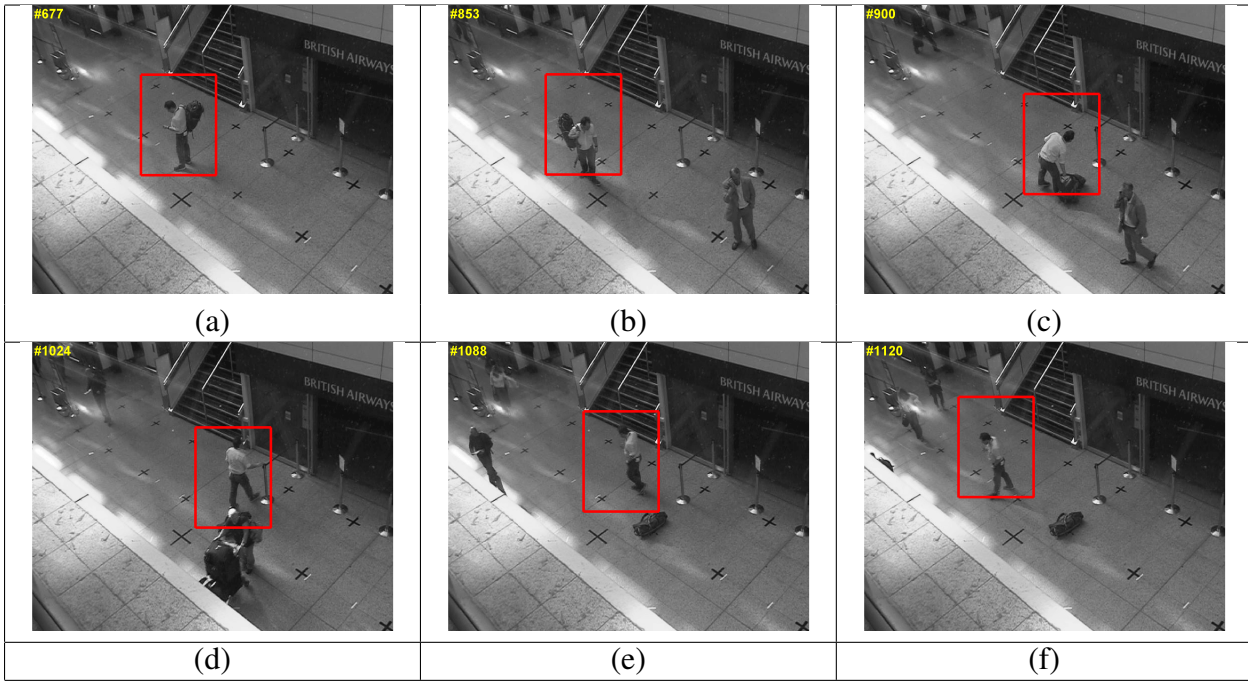


Fig. 5: Detecting sudden-appearance change at another camera